



Bringing Manifold Learning and Dimensionality Reduction to SED Fitters

Shoubaneh Hemmati¹, Peter Capak^{2,3}, Milad Pourrahmani⁴, Hooshang Nayyeri⁴, Daniel Stern¹, Bahram Mobasher⁵, Behnam Darvish⁶, Iary Davidzon², Olivier Ilbert⁷, Daniel Masters¹, and Abtin Shahidi⁵

¹ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA; shemmati@jpl.nasa.gov

² IPAC, California Institute of Technology, 1200 East California Boulevard, Pasadena CA 91125, USA

³ Cosmic Dawn Center (DAWN), University of Copenhagen, Vibenshuset, Lyngbyvej 2, 2100 Copenhagen, Denmark

⁴ University of California, Irvine, Irvine, CA 92697, USA

⁵ University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA

⁶ California Institute of Technology, 1200 E California Boulevard, Pasadena, CA 91125, USA

⁷ Laboratoire D'astrophysique De Marseille, 38 Rue Frédéric Joliot Curie, F-13013 Marseille, France

Received 2019 May 23; revised 2019 July 11; accepted 2019 July 22; published 2019 August 9

Abstract

We show that unsupervised machine learning techniques are a valuable tool for both visualizing and computationally accelerating the estimation of galaxy physical properties from photometric data. As a proof of concept, we use self-organizing maps (SOMs) to visualize a spectral energy distribution (SED) model library in the observed photometry space. The resulting visual maps allow for a better understanding of how the observed data maps to physical properties and allows for better optimization of the model libraries for a given set of observational data. Next, the SOMs are used to estimate the physical parameters of 14,000 $z \sim 1$ galaxies in the COSMOS field and are found to be in agreement with those measured with SED fitting. However, the SOM method is able to estimate the full probability distribution functions for each galaxy up to $\sim 10^6$ times faster than direct model fitting. We conclude by discussing how this acceleration, as well as learning how the galaxy data manifold maps to physical parameter space and visualizing this mapping in lower dimensions, helps overcome other challenges in galaxy formation and evolution.

Key words: galaxies: fundamental parameters – galaxies: statistics

1. Introduction

Understanding how galaxies evolve with time is one of the central questions in observational astronomy. But the physical processes that drive galaxy evolution are a complicated combination of many physical processes that are intertwined. For instance, gas accretion (e.g., Rubin et al. 2012; Somerville & Davé 2015; Zabl et al. 2019) and its cooling into cold molecular gas that forms stars (e.g., Ribaud et al. 2011; Saintonge et al. 2011; Tacconi et al. 2013) is disrupted by feedback from massive stars, active galactic nuclei (AGNs), and/or supernovae (e.g., Springel et al. 2005; Fabian 2012; Tombesi et al. 2015). Mergers of galaxies and other environmental processes further complicate this picture (e.g., Moster et al. 2011; Hopkins et al. 2013; Fensch et al. 2017).

All of these processes are imprinted in the observed spectral energy distributions (SEDs) of galaxies. The shape and normalization of these SEDs encodes physical properties such as the star formation rates and histories, stellar masses, and dust content and distribution of galaxies. To estimate physical properties from the data a detailed physical model is used to synthesize observations for a wide range of possible formation scenarios (e.g., Maraston & Strömbäck 2011; Conroy 2013; Hayward & Smith 2015). This library of synthetic templates is then fitted to the data to determine the distribution of physical processes that might describe a given galaxy. However, the complexity of the physics involved and the high-dimensionality of the physical parameter space mean understanding how the data and model space are related is very difficult. Furthermore, the complexity of the physics means the model grids must be restricted to a controlled range of physical scenarios to make the fitting process computationally feasible.

In Hemmati et al. (2019) and Masters et al. (2019), we studied how self-organizing maps (SOMs; Kohonen 1982) can be

employed to improve redshift measurements. In this paper, we focus our attention toward the benefits of using a manifold learning algorithm (such as SOM) in the measurement of physical parameters of galaxies and present a framework for optimizing the parameters in theoretical models based on the data. In Section 2, we describe the data and methods we adopt for the demonstrations in subsequent sections. Sections 3–5 describe how we use the SOM to carry out parameter estimates and the distinct advantages compared to typical model fitting routines. Finally in Section 6, we discuss ways to build upon this machinery for future large surveys. Throughout this paper all magnitudes are expressed in the AB system (Oke & Gunn 1983) and we use standard cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2. Data and Methodology

The goal of this paper is to demonstrate how unsupervised machine learning can be used to understand and optimize model fitting. To do that we set out a road map based on a demonstration data set. Specific problems will require variations of the method described here. In this section we begin by describing the data we used for the demonstration from the COSMOS survey in Section 2.1 and the model library we have chosen to fit to the data in Section 2.2. We then give a brief overview of the SOM algorithm that we have chosen for this demonstration in Section 2.3 and note some alternative algorithms that could be used.

2.1. The Data

The COSMOS survey (Scoville et al. 2007) offers a suitable data sets for studying galaxy formation and evolution, providing both area ($\sim 2 \text{ deg}^2$) and depth ($\sim 26 \text{ AB mag}$ in the optical) across a broad range of wavelengths with around

two million detected galaxies. We use the broadband photometric data from the latest publicly available multiwavelength catalog in the COSMOS (see Laigle et al. 2016 for details), namely the CFHT/Megacam u^* , Subaru/Suprime-CAM B , V , r , i^+ , z^{++} , VIRCAM/UltraVISTA Y , J , H , K_s , and *Spitzer*/IRAC $ch1$, $ch2$. In this paper, we do not include longer wavelength observations. For demonstration purposes we also fix redshifts to those reported by the COSMOS team and restrict all our analysis to 13,781 galaxies in the redshift range $0.8 < z < 1.2$.

2.2. SED Model Library

We chose the commonly used BC03 simple stellar population synthesis template library (Bruzual & Charlot 2003) restricted to $z \sim 1$ for our demonstration. While not optimal, this simple library is adequate for demonstration purposes. A forthcoming paper, I. Davidzon et al. (2019, in preparation), applies these techniques using more sophisticated hydrodynamical simulations across all redshifts from the Horizon AGN simulation (Laigle et al. 2019).

Here we adopt a Chabrier initial mass function (Chabrier 2003), exponentially declining star formation histories ($e^{-t/\tau}$ with τ values in the range 8.5–10), a subsolar metallicity ($0.4Z_\odot$), a range of stellar ages ($7.7 < \log(\text{age}/\text{yr}) < 10.0$), and a Calzetti et al. (2000) extinction law with a range of reddening values ($0 < E(B - V) < 1$). To generate synthetic photometry that can be fit to our data we redshifted the models and integrated them against the same filter transmissions used for COSMOS observations (see Section 2.1). This yields 13,776 models with the same parameters as the observed data.

2.3. Unsupervised Learning Algorithm

We chose to use the SOM algorithm for this demonstration because it provides an easily understood visual map of the high-dimensional data along with a vector representation of the model space. Furthermore, we have used it in previous papers, Hemmati et al. (2019) and Masters et al. (2019), for related problems so contemporary readers would be familiar with the method. We note that other manifold learning and dimensionality reduction methods such as but not limited to the t-SNE (t-Distributed Stochastic Neighbor Embedding; van der Maaten & Hinton 2008), UMAP (Uniform Manifold Adaptation and Projection; McInnes et al. 2018), and GNG (Growing Neural Gas networks; Fritzke 1994) can also be employed and may be optimal for other problems.

SOMs are a class of unsupervised neural networks that reduce the dimensionality of data, while preserving the topology. Therefore, neighboring points in the N -dimensional space are going to be neighbors in the 2D projected map. Each neuron corresponds to a pixel on the map as well as a vector into the photometric data space. The neurons/vectors can be linearly combined to represent any object in the data set. Alternatively, the neurons can be thought of as bins in the high-dimensional data space. Data points that are statistically similar will get assigned to the same bin, reducing the number of point required to describe the data set. Other manifold learning techniques (e.g., t-SNE, UMAP) have the advantage of preserving not just the topology of the higher dimensional space but also the distance. However, they lack the binning aspect of the SOM algorithm which makes them less favorable for our purpose.

In previous works (e.g., Masters et al. 2015; Hemmati et al. 2019; Masters et al. 2019), we demonstrated how the manifold of galaxy colors can be learned using an SOM. We then showed that an SOM can be used to understand and calibrate photometric redshifts as well as understand variations in galaxy spectral types with color. We refer the interested reader to those works for more details about the methodology.

For this work we use the SOMPY package, a python library for SOMs and adopt an 80×60 rectangular grid. The size is chosen to represent the data distribution well (Hemmati et al. 2019). The consecutive colors of galaxies as measured in the filters listed in Section 2.1 define the multidimensional (11D) color-space to be learned. Hence, each cell on the constructed grid has an 11-dimensional weight vector whose values (initially assigned by principle component analysis) change until they fully represent the training data. In the SOM algorithm, the weights of the neurons get updated by the following equation:

$$W_i(t+1) = W_i(t) + \alpha(t)h(t)([x(t) - w_i(t)]), \quad (1)$$

where t is the current iteration, α and h are the learning rate and the neighborhood distance, respectively, and x is the training data point. The distance between the data points and the weight vectors at each step is chosen to be the Euclidean distance.

3. Visualizing and Optimizing Model Libraries

Having a large number of physical parameters (dimensions) in a model SED library is valuable for understanding the galaxy population. However, with large numbers of parameters traditional model fitting analysis becomes computationally unfeasible even with advanced techniques such as Monte Carlo Markov Chain (MCMC) fitting (e.g., Johnson et al. 2013; Speagle et al. 2016; Wilkinson et al. 2017). Furthermore, understanding how each parameter is shaping the overall fit, which ones are correlated, and what is actually constrained by the data becomes very challenging.

A primary advantage of using unsupervised dimensionality reduction methods like the SOMs is the ability to visualize complex data spaces. In the model space, the dimensionality increases with the number of physical properties considered, so varying the IMF or adding different dust properties would increase the dimensionality. In this paper we focus on visualizing the models in data space, but note that the same techniques could be used to visualize in the model parameter space. For photometric data, the dimensions of the observed parameter space increase with increased number of colors (filters). In other words, increasing the wavelength range by adding broadband filters or the resolution by adding intermediate band filters would increase the dimensionality of the data space.

We begin by taking the synthetic model photometry described in Section 2.2 and creating an SOM that represents those synthetic photometry data. In Figure 1 we visualize model parameters in photometric data space by coloring each pixel in the SOM with its median physical parameter from the models. The smooth areas of the maps represent regions where the data space directly maps to physical parameters, so the data can strongly constrain the physical parameter in question. The more stochastic parts of the map indicate regions where the data are not constraining of a given physical parameter, so the same data can describe multiple possible physical properties. It is clear from looking at these maps that there is a region

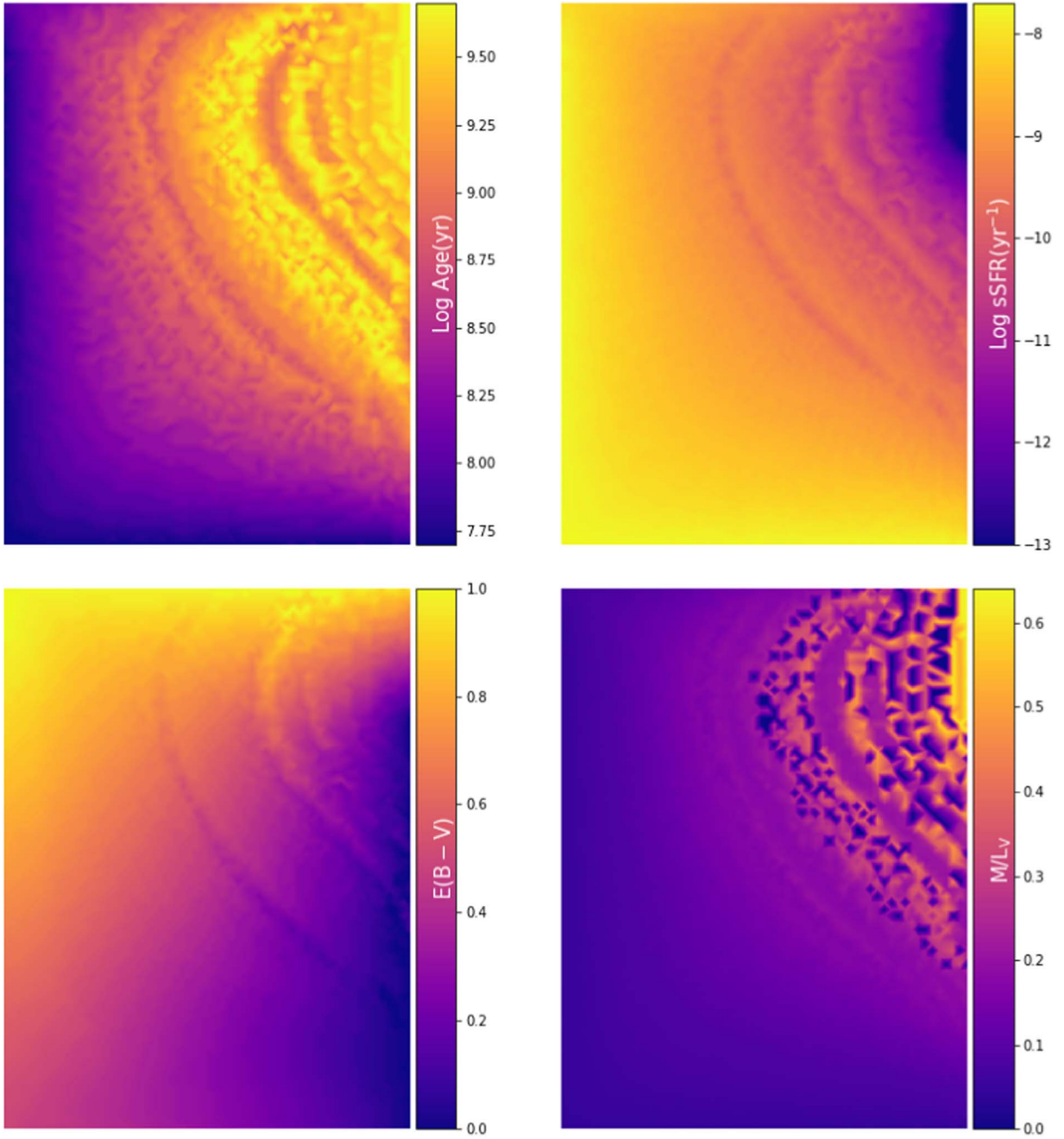


Figure 1. Composite stellar population SEDs from our BC03 model library visualized on an 80×60 SOM grid. The consecutive colors of the model SEDs in COSMOS optical and near-infrared broadband filter sets ($u - B$, $B - V$, $V - r$, $r - i^+$, $i^+ - z^{++}$, $z^{++} - Y$, $Y - J$, $J - H$, $H - K_s$, $K_s - ch1$, $ch1 - ch2$) are used to train the SOM. The four maps here show linearly interpolated medians of the physical properties of the models mapped to each cell. From top left to bottom right, colors correspond to stellar age, specific star formation rate (sSFR), dust reddening, and mass to light ratio.

of data space where age and mass-to-light ratios are poorly constrained by the data and other regions where the properties are well constrained.

The next question is which portions of the model space are occupied by real galaxies and do the galaxies occupy smooth or stochastic parts of the model space. In Figure 2 we map the data for COSMOS $z \sim 1$ from Section 2.1 galaxies to the model SOM. The empty regions in the map indicate regions where no real object corresponds to those model parameters. For instance, the large empty region in the bottom right of the plot indicates that a combination of very young age, low extinction, and high star formation rate per unit mass (specific

star formation rate, and sSFR) is not observed in COSMOS at $z \sim 1$. Furthermore, the points at the very edge of the plot are also at the edge of the model space and hence are not well fit. The presence of these points indicates galaxies that are not well described by the models and that additional parameters, such as the presence of a AGNs may be needed.

This or a similar analysis can be used to optimize the model library. Regions of the model space that are not occupied by real objects should be removed or down weighted based on the data statistics. Including them in a fitting library with a flat prior will lead to biased physical property estimates because noise in the data will scatter objects into regions of model space

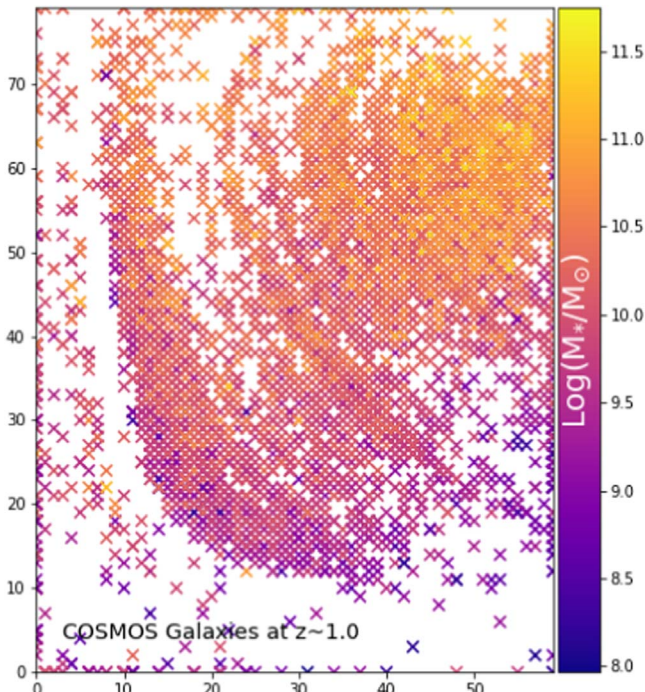


Figure 2. COSMOS galaxies at $z \sim 1.0$ are mapped to the SOM trained with theoretical models. Empty regions of the SOM represent combinations of the theoretical parameter space that do not have observed counterparts. Objects at the edge of the map typically are poorly fit by the model. This visualization enables the optimization of model libraries based on information from the observed data space. Galaxies in the map are color-coded with their estimated stellar masses from Laigle et al. (2016).

that do not physically exist. Furthermore, the expansion of the model library could be guided by the points that are not well described by the model library.

4. Optimizing Model Fitting

Mapping an observed galaxy to the SOM trained on the model libraries is mathematically equivalent to finding the best-fit template in a model library. Here we used a module called “project-data” from the *SOMPY* package, which uses the k -nearest neighbors algorithm. We note that the physical property maps as seen in Figure 1 are produced by interpolating the median of the physical properties of models mapped to the cells of the grid. However, because the SOM optimally represents the data space, observationally indistinguishable models have already been assigned to the same SOM vector. This methodology effectively “compresses” the computation by pregrouping models that will have indistinguishable fits to the data. So, rather than fitting every model one only needs to fit each point in the SOM to estimate the posterior likelihood distribution.

In our simple demonstration (see Figure 4) the likelihood distribution in physical parameter space is calculated by assigning the median property of the models at each SOM point. We then weight each SOM point by its likelihood with respect to the data in the same way as if one were fitting a model library. We note that a better weighting scheme could easily be devised, but is beyond the scope of the demonstration here.

In Figure 3, we compare best-fit stellar mass measured from the SOM to that from the COSMOS catalog estimated using the LePhare SED fitting code (Arnouts et al. 1999; Ilbert et al. 2006).

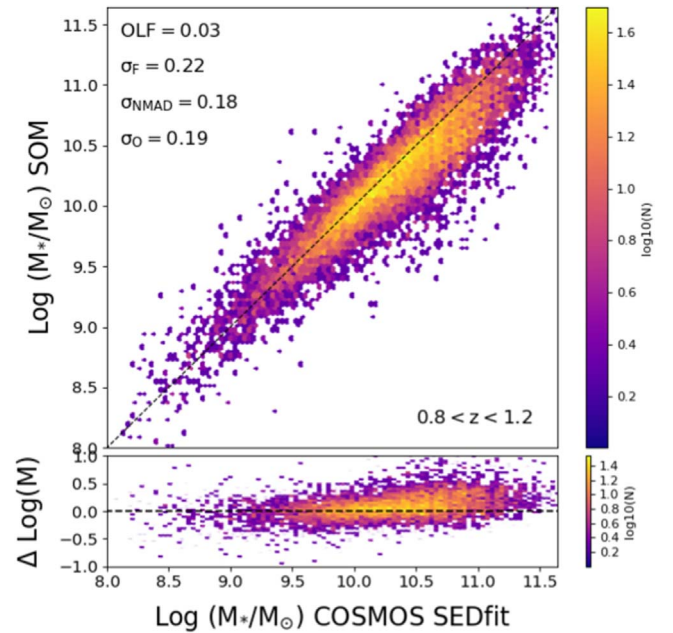


Figure 3. Stellar masses of COSMOS $z \sim 1$ galaxies measured by mapping them to the SOM trained with BC03 models (SOMfit) compared to stellar masses from the Laigle et al. (2016) catalog measured with LePhare. The differences are well within the expectations from using different model libraries.

In this comparison, we use statistics similar to those defined in Nayyeri et al. (2017), where $\sigma_F = \text{rms}(\log(M_{\text{SOM}}) - \log(M_{\text{COSMOS}}))$, $\sigma_{\text{NMAD}} = 1.48 \times \text{median}(|\log(M_{\text{SOM}}) - \log(M_{\text{COSMOS}})|)$, and $\sigma_O = \text{rms}(\log(M_{\text{SOM}}) - \log(M_{\text{COSMOS}}))$ after removing the outliers with the outlier fraction (OLF) defined as the fraction of objects with $\Delta \log(M) = (\log(M_{\text{SOM}}) - \log(M_{\text{COSMOS}})) > 0.5$. The scatter of ~ 0.2 and 3% OLF is comparable with the values reported in the literature when various SED fitting codes are compared (e.g., Mobasher et al. 2015; Nayyeri et al. 2017) and we note more advanced models including emission lines and other features (i.e., delayed SFH and two attenuation curves) were included in the COSMOS results. Furthermore, using a more thorough weighting scheme and including redshift errors would likely further improve results.

Another compelling feature of the SOM is that the errors on the physical properties are quantified. The method naturally builds up a probability distribution function for each physical parameter based on the likelihoods of the models. For instance, one can perturb the photometric error in each broadband filter to have many realizations of a galaxy SED and mapping them to the SOM. In Figure 4 we show the kernel density function of 1000 realizations of the SED, drawn from the estimated errors, mapped to the SOM with contours for two sample galaxies. The corresponding distribution of physical properties are plotted in the inset panels.

It is clear in this visualization how the error contour for a given galaxy spreads over the SOM grid and how that is related to the physical parameters shown in Figure 1.

By comparing Figures 4 and 1, it is clear how parameters are covariant with error. In many SED fitting routines, all theoretical model SEDs have the same probability of being fitted to a set of observations. A comparison of Figures 2 and 1 shows that photometric errors will create probability in nonphysical regions of the model space where no objects occur. So the use of a flat prior creates a clearly biased result. A

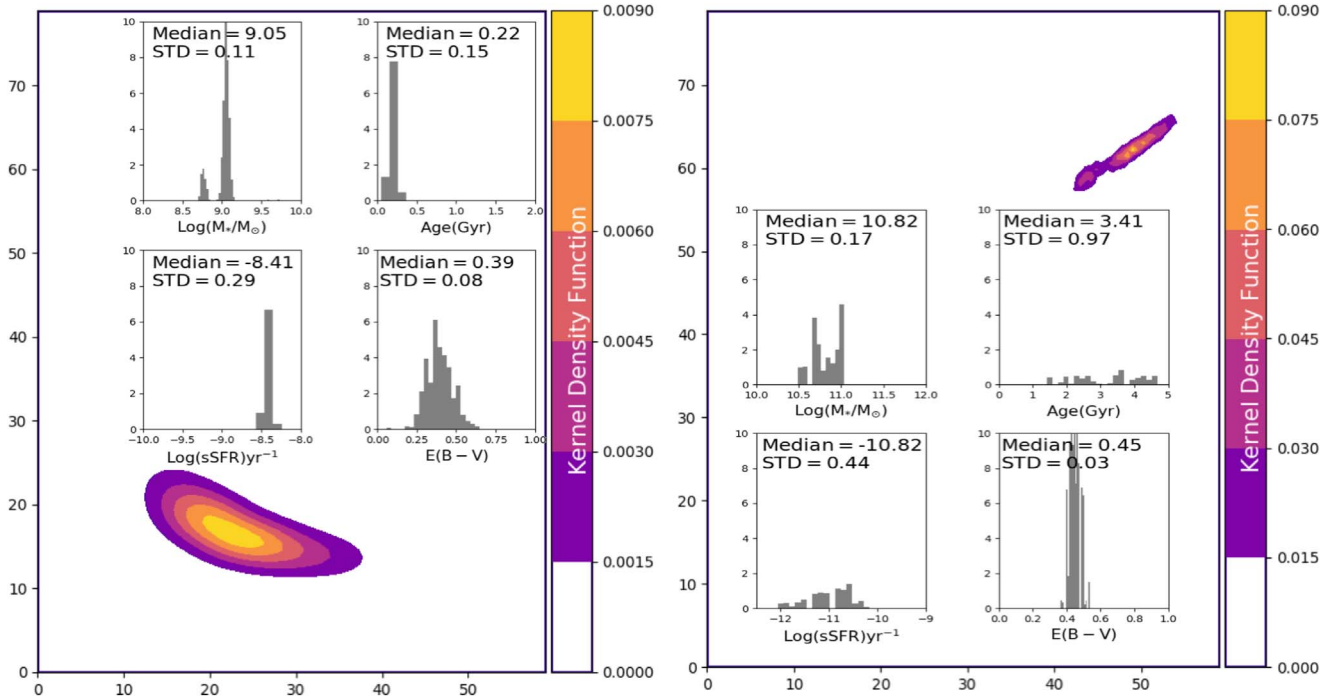


Figure 4. To quantify the uncertainties in the measurement of physical parameters with the SOM, many realizations (here 1000) of two sample galaxy SEDs perturbed within the photometric errors in each band are mapped to the SOM to generate the kernel density function. The inset plots represent the confidence on each of the measured physical parameters (top left to bottom right panels: stellar mass, stellar age, sSFR, and $E(B - V)$).

better approach would be to use the density of data on the SOM grid as a weight on the model fit. As demonstrated in Masters et al. (2019) this results in a less biased result while still preserving the ability to fit legitimately rare sources.

5. Computational Speed

Measuring galaxy physical properties using SOMs is also computationally efficient. Once the SOM is trained with a library of model SEDs, there is no need to keep the large number of models in the memory and any set of observations can be mapped to the trained SOM rapidly. In addition, the size of the SOM is determined by the data space, not the model space, so the required computational time to fit a set of models is largely invariant to the complexity and size of model grids and is instead determined by the dimensionality of the data set. Thus it is possible to fit multiple very complex models to the data. This is a significant advantage for future large area surveys with billions of galaxies such as *Euclid*, LSST, and *WFIRST* because it will allow for rapid refitting of multiple SED model libraries.

Furthermore, the mapping of galaxies to the SOM is inherently parallel and hence is easily applicable on a GPU or a large cluster. To verify the speed up we converted the mapping phase (i.e., finding the best matching unit of the SOM with least χ^2) to a vectorized GPU friendly code using *PyTorch* (Paszke et al. 2017) and *Numba* (Lam et al. 2015). Figure 5 shows the processing time of measuring physical properties of galaxies using the LePhare SED fitting code (REF) compared to mapping the galaxies onto the SOM using both a CPU on a Core-i7 machine and an *Nvidia Titan V GPU*. The SOM is an ~ 3 –4 orders of magnitude faster than LePhare on a CPU and ~ 5 –6 orders of magnitude faster on a GPU. This means even a single threaded CPU could measure the physical parameters of

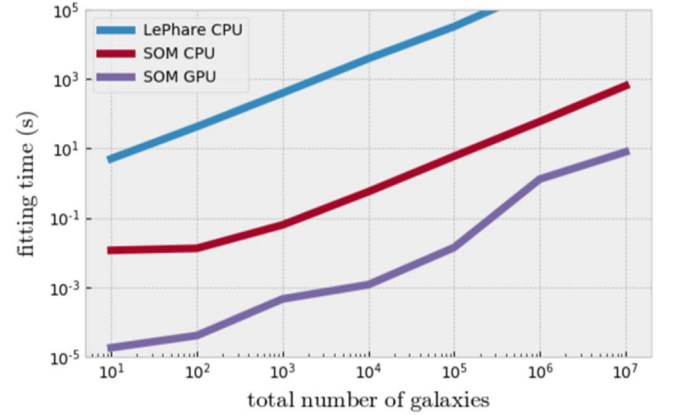


Figure 5. Computational speed of mapping different numbers of galaxies to the SOM (on a CPU and on a GPU) compared to SED fitting with LePhare. Using our mapping we increase the performance by $\sim 10^{-4}$ on a CPU and $\sim 10^{-6}$ on a GPU compared to SED fitting.

over a million galaxies per minute and a single GPU could recompute stellar masses for the 30 billion objects expected from LSST in under nine hours compared with several years for LePhare. So even a modest cluster of GPU machines could recompute these parameters for all of LSST interactively.

6. Discussion

We present the idea of using manifold learning as an augmentation to traditional SED fitting. We present the SOM method in particular as a way of visualizing and optimizing a library of theoretical models. We then show that the same method can be used for fast estimation of galaxy physical properties and how the data maps to these properties.



In previous works we showed that the manifold mapping approach improves photometric redshifts (e.g., Masters et al. 2015, 2017; Hemmati et al. 2019; Masters et al. 2019). In this work, as a proof of concept, we constrained our analysis to only a small redshift bin and a simple library of model SEDs to clearly demonstrate the method. We show that the visualization and computational improvements have clear utility for other studies.





Beyond fitting objects, the techniques we present here have multiple uses. Most popular galaxy classifications in extragalactic astronomy use only a few broadband colors to select distinct classes of galaxies such as the Lyman break, Balmer break, or the “UVJ” selections (e.g., Steidel et al. 1999; Williams et al. 2009; Nayyeri et al. 2014). Manifold mapping techniques can offer an excellent alternative tool that uses all of the color information. For instance, the SOM, with its preserved topology could be used to specify N -dimensional color selections for specific classes of galaxies. The exact cuts could, for instance, be defined using the visualization presented in Figure 1. Frontera-Pons et al. (2017) using another manifold learning technique (i.e., denoising autoencoders) on COSMOS rest-frame colors, clearly demonstrated the benefit of using all information in the SED for classification purposes. The potential of such generative models (e.g., variational autoencoders, generative adversarial networks), which have recently had huge success in other fields, goes beyond classifications in astronomy. As also noted by Frontera-Pons et al. (2017), these models once trained on noisy data can decode and generate the denoised distribution of data. This feature is of particular interest when observed (noisy) data space is to be compared/fitted with the theoretical model space.

The speed of the method discussed in this paper also allows for new possibilities. For instance, it could be applied to entire multicolor data sets at the pixel or resolution element level. This would enable exploration of resolved kiloparsec-scale distribution of physical properties (e.g., Wuyts et al. 2012; Hemmati et al. 2014, 2015) or definition of object deblending based on these maps. Finally, other observational data besides colors could be included in this method to create multimeasurement studies or fits. For instance, environmental measures or morphology could be included. We are now extending this work to more complex models based on hydrodynamical simulations (I. Davidzon et al. 2019, in preparation) and use a similar technique to predict higher resolution spectral features like emission line strengths from broadband information (S. Hemmati et al. 2019, in preparation).

We wish to thank the referee for very helpful comments that improved the content and presentation of this paper. This work used SOMPY, a python package for SOMs (main contributors: Vahid Moosavi @sevamoo, Sebastian Packmann @sebastian-dev, Iván Vallás @ivallesp). We are thankful to NVIDIA for the GPU granted as their academic grant program. Parts of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

ORCID iDs

Shoubaneh Hemmati  <https://orcid.org/0000-0003-2226-5395>
Peter Capak  <https://orcid.org/0000-0003-3578-6843>

Milad Pourrahmani  <https://orcid.org/0000-0003-3351-5986>
Hooshang Nayyeri  <https://orcid.org/0000-0001-8242-9983>
Daniel Stern  <https://orcid.org/0000-0003-2686-9241>
Behnam Darvish  <https://orcid.org/0000-0003-4919-9017>
Iary Davidzon  <https://orcid.org/0000-0002-2951-7519>
Olivier Ilbert  <https://orcid.org/0000-0002-7303-4397>
Daniel Masters  <https://orcid.org/0000-0001-5382-6138>

References

- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
Chabrier, G. 2003, *PASP*, 115, 763
Conroy, C. 2013, *ARA&A*, 51, 393
Fabian, A. C. 2012, *ARA&A*, 50, 455
Fensch, J., Renaud, F., Bournaud, F., et al. 2017, *MNRAS*, 465, 1934
Fritzke, B. 1994, in Proc. 7th Int. Conf. on Neural Information Processing Systems, NIPS’94 (Cambridge, MA: MIT Press), 625
Frontera-Pons, J., Sureau, F., Bobin, J., & Le Floch, E. 2017, *A&A*, 603, A60
Hayward, C. C., & Smith, D. J. B. 2015, *MNRAS*, 446, 1512
Hemmati, S., Capak, P., Masters, D., et al. 2019, *ApJ*, 877, 117
Hemmati, S., Miller, S. H., Mobasher, B., et al. 2014, *ApJ*, 797, 108
Hemmati, S., Mobasher, B., Darvish, B., et al. 2015, *ApJ*, 814, 46
Hopkins, P. F., Cox, T. J., Hernquist, L., et al. 2013, *MNRAS*, 430, 1901
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841
Johnson, S. P., Wilson, G. W., Tang, Y., & Scott, K. S. 2013, *MNRAS*, 436, 2535
Kohonen, T. 1982, *Biol. Cybern.*, 43, 59
Laigle, C., Davidzon, I., Ilbert, O., et al. 2019, *MNRAS*, 486, 5104
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
Lam, S. K., Pitrou, A., & Seibert, S. 2015, in Proc. Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM’15 (New York, NY: ACM), 7:1
Maraston, C., & Strömback, G. 2011, *MNRAS*, 418, 2785
Masters, D., Capak, P., Stern, D., et al. 2015, *ApJ*, 813, 53
Masters, D. C., Stern, D., Cohen, J. G., et al. 2017, *ApJ*, 841, 111
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, arXiv:1904.06394
McInnes, L., Healy, J., & Melville, J. 2018, arXiv:1802.03426
Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, *ApJ*, 808, 101
Moster, B. P., Macciò, A. V., Somerville, R. S., Naab, T., & Cox, T. J. 2011, *MNRAS*, 415, 3750
Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, *ApJS*, 228, 7
Nayyeri, H., Mobasher, B., Hemmati, S., et al. 2014, *ApJ*, 794, 68
Oke, J. B., & Gunn, J. E. 1983, *ApJ*, 266, 713
Paszke, A., Gross, S., Chintala, S., et al. 2017, NIPS-W, version 0.3
Ribaudo, J., Lehner, N., Howk, J. C., et al. 2011, *ApJ*, 743, 207
Rubin, K. H. R., Prochaska, J. X., Koo, D. C., & Phillips, A. C. 2012, *ApJL*, 747, L26
Saintonge, A., Kauffmann, G., Kramer, C., et al. 2011, *MNRAS*, 415, 32
Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
Somerville, R. S., & Davé, R. 2015, *ARA&A*, 53, 51
Speagle, J. S., Capak, P. L., Eisenstein, D. J., Masters, D. C., & Steinhart, C. L. 2016, *MNRAS*, 461, 3432
Springel, V., Di Matteo, T., & Hernquist, L. 2005, *MNRAS*, 361, 776
Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, *ApJ*, 519, 1
Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, *ApJ*, 768, 74
Tombesi, F., Meléndez, M., Veilleux, S., et al. 2015, *Natur*, 519, 436
van der Maaten, L., & Hinton, G. 2008, Journal of Machine Learning Research, 9, 85
Wilkinson, D. M., Maraston, C., Goddard, D., Thomas, D., & Parikh, T. 2017, *MNRAS*, 472, 4297
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, *ApJ*, 691, 1879
Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, *ApJ*, 753, 114
Zabl, J., Bouché, N. F., Schroetter, I., et al. 2019, *MNRAS*, 485, 1961